| Proposed By | Henry Chan, Ken Lunde, Michel Suignard |
|---|---|
| Title | Re-mapping of H-9B4C and HB0-A2CD (kIRG_HSource) |
| Meeting | IRG #44 |
| Status | Individual Contribution |
| Action | For Consideration by the IRG |
| | (Discussion primer: IRGN 2074) |
| Pages | 3 |

Hong Kong is planning to issue the Hong Kong Character Set (HKCS) in 2015 to replace the existing Hong Kong Supplementary Character Set (HKSCS) standard and migrate away from dependencies on Big-5. However, several mapping issues have been identified.

**Issue #1**

The form for H-9B4C (see Figure 1.1) specified in the HKSCS standard uses the modern variant of 既 as its top component, which is different from most of the other forms mapped to U+69E9 (see Figure 1.2). The form used would more appropriately be mapped to U+3BA3 (see Figure 1.3).



Figure 1.1 – Form for H-9B4C, currently mapped to U+69E9



Figure 1.2 – ISO/IEC 10646 Code Chart for U+69E9



Figure 1.3 – ISO/IEC 10646 Code Chart for U+3BA3

Compatibility is not an issue. Currently the default fonts for the Windows and OS X operating systems in use in Hong Kong display the T-Source form for U+69E9. Choosing the character 㮣 in an input method editor would result in U+3BA3. Furthermore, as the Hong Kong SAR government strongly encourages the use of ISO/IEC 10646 in computer systems, round-trip mapping would not be a problem.

When the Hong Kong Character Set 2015 (HKCS) is released, which will comprehensively cover characters in use in Hong Kong, computer vendors may finally release fonts that follow the relevant glyph standards. However, if that happens, the glyph for characters encoded using U+69E9 would change unexpectedly. Since U+69E9 and U+3BA3 would display as the same glyph, users would also be unable to produce the traditional variant, which is not preferred by the forthcoming HKCS standard.

The Chinese Language Interface Advisory Committee (CLIAC) of the Office of the Chief Information Officer, HKSARG has discussed changing the glyph at U+69E9/H-9B4C to use the traditional variant of 既 as its top component, similar to the T-Source of U+69E9. However, it was concluded that the glyph shape for H-9B4C could not be modified. Therefore, the only remaining option is to re-map H-9B4C to U+3BA3.

**Re-map Precedent**

Corrections to HKSCS mappings have been done in the past. Per WG2 N4621, the H-Source for U+3D1D was changed to U+2A3ED, to better match the glyphs for other regions. The proposal was accepted by WG2 per WG2 N4604 Recommendation M63.05.

**Recommendation**

Re-map H-9B4C to U+3BA3.

**Issue #2**

HB0-A2CD is currently mapped to U+5344. However, in Big-5, 0xA2CD is grouped with other Suzhou Numerals. It should instead map to U+3039 HANGZHOU NUMERAL TWENTY instead. Also, the currently unmapped characters 0xA2CC and 0xA2CE should also be mapped to U+3038 and U+303A, respectively.

The mapping omission has likely arisen due to U+3038 HANGZHOU NUMERAL TEN, U+3039 HANGZHOU NUMERAL TWENTY, and U+303A HANGZHOU NUMERAL THIRTY being added to the standard at a later point in time (ISO/IEC 10646-1:2000).

The T-Source for U+5344 has been updated from T1-243F to T3-2138. T1-243F corresponds to Big-5 code point 0xA2CD.

**Summary**

| Big-5 Code Point | Current UCS Mapping | Recommended UCS Mapping |
|---|---|---|
| 0xA2C3 丨 | U+3021 HANGZHOU NUMERAL ONE | NO CHANGE |
| 0xA2C4 刂 | U+3022 HANGZHOU NUMERAL TWO | NO CHANGE |
| … | … | … |
| 0xA2CA 亖 | U+3028 HANGZHOU NUMERAL EIGHT | NO CHANGE |
| 0xA2CB 夂 | U+3029 HANGZHOU NUMERAL NINE | NO CHANGE |
| 0xA2CC 廿 | U+FFFD (Unmapped) | U+3038 HANGZHOU NUMERAL TEN |
| 0xA2CD 卄 | U+5344 CJK UNIFIED IDEOGRAPH-5344 | U+3039 HANGZHOU NUMERAL TWENTY |
| 0xA2CE 卅 | U+FFFD (Unmapped) | U+303A HANGZHOU NUMERAL THIRTY |

**Suggestion**

Re-map HB0-A2CD to U+3039, map HB-A2CC to U+3038, and map HB-A2CE to U+30 for backward compatibility, some implementers may choose to map both code points (U+3039 and U+5344) to the same glyph.

- End of document

Date : 2015 - 08 - 25

ISO/IEC JTC1/SC2/WG2/IRG

Ideographic Rapporteur Group

(IRG)

Source/Contribution Identifier :   Hong Kong Special Administrative Region, China (HKSAR)

Meeting :   44<sup>th</sup> IRG Meeting in Beijing, China

Title :   Feedback to IRGN2077 on the Re-mapping of H-9B4C and HB0-A2CD (kIRG_Hsource)

Status :   Member feedback for discussion

**Proposal (1): Remapping of H-9B4C from U+69E9 to U+3BA3**

2.	The table below shows the glyphs of U+69E9 and U+3BA3 in the ISO/IEC 10646-1:1993 and later versions:

| ISO/IEC 10646-1:1993 | 105/233 **69E9** E-2E65 3-4C64 E-1469 3-4468 1-576B 1-5575 [Unification of the glyphs 櫢 and 櫢] |
| --- | --- |
| ISO/IEC 10646-1:2000 | 105/233 **69E9** E-2E65 3-4C64 E-1469 3-4468 1-576B 1-5575  059/163 **3BA3** 3-4167 3-3371 [Duplication of the glyph 櫢] |
| ISO/IEC 10646:2003 | 105/233 **69E9** E-2E65 3-4C64 E-1469 3-4468 1-576B 1-5575  059/163 **3BA3** 3-4167 3-3371 |
| ISO/IEC 10646:2011 | **69E9** 木 75.10 GE-2E65 T3-4C64 J3-7623 K1-576B V2-8D3A H-9B4C  **3BA3** 木 75.9 GHZ-21257.01 T3-4167 |
| ISO/IEC 10646:2012 | **69E9** 木 75.10 GE-2E65 H-9B4C T3-4C64 J3-7623 K1-576B V2-8D3A  **3BA3** 木 75.9 GHZ-21257.01 T3-4167 |
| ISO/IEC 10646:2014(E) | **69E9** 木 75.10 GE-2E65 H-9B4C T3-4C64 J3-7623 K1-576B V2-8D3A  **3BA3** 木 75.9 GHZ-21257.01 T3-4167 |

3.　　It is evident that the glyphs 槑 and 槑 have been unified under U+69E9 in the BMP since ISO/IEC 10646-1:1993.　The coding of T3-4167 as U+3BA3 in Extension A in ISO/IEC 10646-1:2000 breached the unification rule and resulted in the duplicated coding of the glyph 槑.

4.　　The duplication is not desirable, but the HKSAR cannot accept the remapping proposal due to the following reasons:

a.　The mapping of H-9B4C 槑 (H-9B4C) to U+69E9 conforms with the unification rule and is a legitimate one.　The glyph was linked to U+69E9 when it was first published in the HKSCS -1999: 槑 (69E9 / 69E9).

b.　The remapping of H-9B4C cannot rectify the problem of duplication so long as GE-2E65 槑 (GE-2E65) remains in U+69E9.　China may consider remapping or removing GE-2E65.　The HKSAR however finds it important to uphold the IRG's unification principle and not to disunify properly unified glyphs simply because a duplicate crops up.

There are other glyphs with the component "既" or "旣" in the BMP.　Some are unified while some are coded separately, as shown in the table below.

| Separately-coded unifiable glyphs | 66A8 日 72.10 | 暨 G0-745F | 暨 HB1-BA5B | 暨 T1-6924 | | | |
| | 66C1 日 72.12 | 暨 GE-2C7B | | 暨 T3-5123 | 暨 J0-5A7A | 暨 K1-5B70 | |
| | 6982 木 75.9 | 概 G0-3845 | 概 HB1-B7A7 | 概 T1-644D | 概 J0-3335 | | |
| | 69EA 木 75.11 | 槩 GE-2E66 | | 槩 T3-4C60 | 概 J3-7624 | 概 K0-4B48 | |
| | 6E89 水 85.9 | 溉 G0-3848 | 溉 HB1-B540 | 溉 T1-6046 | | | |
| | 6F11 水 85.11 | 漑 GE-3071 | | 漑 T3-4764 | 漑 J0-5E74 | 漑 K0-4B49 | |
| Unified glyphs | 5605 口 30.11 | 嘅 GH-1231 | 嘅 H-9DEF | 嘅 T3-4636 | 嘅 J1-362A | 嘅 K2-284A | 嘅 V2-713E |
| | 5ED0 广 53.11 | 廐 GE-2934 | 廐 H-9DFB | 廐 T3-4660 | 廐 J0-567D | 廐 K0-4F2A | |

| | 6168<br>心 61.9 | 慨<br>G0-3F2E | 慨<br>HB1-B46E | 慨<br>T1-5F35 | 慨<br>J0-3334 | 慨<br>K0-4B46 | 慨<br>V1-5577 |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | 6461<br>手 64.11 | 摡<br>G3-4A76 | 摡<br>HB2-D962 | 摡<br>T2-3C27 | 摡<br>J1-4065 | 摡<br>K2-3665 | 摡<br>V2-8C5A |
| | 8507<br>艸 140.11 | 蕑<br>G3-6A28 | 蕑<br>HB2-DFBD | 蕑<br>T2-4662 | 蕑<br>J1-592F | 蕑<br>K2-5940 | |

Our concern is that the disunification of 槩 and 槩 will set a precedent for handling duplication. What if duplicates of other properly unified glyphs, say U+5605, U+5ED0, U+6168, U+6461 and U+8507, are found in other extensions? Are all member bodies prepared to accept disunification and remap their glyphs?

c. Remapping inevitably leads to the round-trip compatibility problem. H-9B4C may not be a frequently used character, but its users in Hong Kong are bound to suffer. The CJK block has long been supported by popular platforms and is widely adopted in software and user data across the world. The impact of disunification and the compatibility issue should not be underestimated as data, software and systems outside Hong Kong are involved.

**Proposal (2): Addition/revision of Big5 symbol mappings**

5. It is proposed in IRGN2077 that three Big5 symbols, namely "十"(0xA2CC), "卄"(0xA2CD) and "卅"(0xA2CE), be mapped/remapped to U+3038, U+3039 and U+303A. We find the proposal reasonable and support the addition/revision of the mappings.

6. As shown in the table below, these three symbols are included in the Symbol Block of the Big5 standard and three similar glyphs are coded as 1xA451, 1xA47B and 1xA4CA in the Character Block.

| Symbol Block<br>(A140-A2CE) | A2C0  Ⅷ  Ⅸ  Ⅹ  │  ╎  ╏  乂   furnace  ⚋  ⚌  ⚍  文  十  卄  卅  A |
| --- | --- |
| Character Block<br>(A440-C67E) | A450  匕  十  卜  乂  三  下  丈  上  丫  九  凡  久  么  也  乞  于<br>A460  亡  兀  刃  勹  千  叉  口  土  士  夕  大  女  子  孑  孓  寸<br>A470  小  尢  尸  山  川  工  己  已  巳  巾  干  卅  弋  弓  才<br>A4A0     丑  丐  不  中  丰  丹  之  尹  予  云  井  互  五  亢  仁<br>A4B0  什  仃  仆  仇  仍  今  介  仄  元  允  內  六  兮  公  冗  凶<br>A4C0  分  切  刈  勻  勾  勿  化  匹  午  升  卅  卞  厄  友  及  反 |

4

7.      In establishing association between symbols and characters, the ISO/IEC 10646 associates the additional Suzhou numeral "卄"(U+3039) with the character "卄" (U+5344) despite the fact that 1xA47B has been mapped to "卝"(U+5EFE) instead. We suggest that the IRG consider changing the association from U+3039 ≈ U+5344 to U+3039 ≈ U+5EFE.

**Additional Suzhou numerals**

| | | |
|---|---|---|
| 3038 | 十 | HANGZHOU NUMERAL TEN |
| | | ≈ 5341 十 |
| 3039 | 卄 | HANGZHOU NUMERAL TWENTY |
| | | ≈ 5344 卄 |
| 303A | 卅 | HANGZHOU NUMERAL THIRTY |
| | | ≈ 5345 卅 |

| 5344 — 1.2 | 卄 G5-3022 | 卄 HB0-A2CD | 卄 T3-2138 | | 卄 K0-6C7E |
|---|---|---|---|---|---|
| 5EFE 卝 55.0 | 卝 G0-5E43 | 卝 HB1-A47B | 卝 T1-445C | 卝 J0-5730 | 卝 K2-3079 | 卝 V1-5441 |

**Conclusion**

8.      We cannot accept Proposal (1).   As a longstanding member body of the IRG, the HKSAR has the responsibility to uphold the unification principle to ensure a stable and consistent standardisation.   Disunifying the glyphs will set a bad precedent and adversely affect the integrity and compatibility of user data.   We urge the IRG to keep H-9B4C in U+69E9.

9.      We support the addition/revision of the mappings as put forward in Proposal (2). To tidy things up, we suggest further changing the association U+3039 ≈ U+5344 to U+3039 ≈ U+5EFE.

End of document